Discussion of “The Shrinkage-Delinkage Trade-off”

Restricted variational families (mean field) can lead to poor posterior approximations.

Two very common responses to this problem are:

- “Machine learning”: Ignore it (evaluate using some other criteria, like prediction)
- “Modern VI”: Use more expressive families (at a computational cost)

This paper tries to understand *how* the restricted family goes wrong. *fire emoji*

Discussion of “The Shrinkage-Delinkage Trade-off”

Restricted variational families (mean field) can lead to poor posterior approximations.

Two very common responses to this problem are:

- “Machine learning”: Ignore it (evaluate using some other criteria, like prediction)
- “Modern VI”: Use more expressive families (at a computational cost)

This paper tries to understand *how* the restricted family goes wrong. *fire emoji*

This paper’s most (initially) surprising conclusion is probably this:

Theorem 3.6, paraphrased

Let the target distribution have the constant ε-correlation matrix.
As the dimension n of the matrix goes to infinity:

- Each marginal mean field variance is wrong by ε
- The per-component entropy gap $\to 0$

Discussion of “The Shrinkage-Delinkage Trade-off”

Restricted variational families (mean field) can lead to poor posterior approximations.

Two very common responses to this problem are:

- “Machine learning”: Ignore it (evaluate using some other criteria, like prediction)
- “Modern VI”: Use more expressive families (at a computational cost)

This paper tries to understand *how* the restricted family goes wrong. *fire emoji*

This paper’s most (initially) surprising conclusion is probably this:

Theorem 3.6, paraphrased

Let the target distribution have the constant ε-correlation matrix.
As the dimension n of the matrix goes to infinity:

- Each marginal mean field variance is wrong by ε
- The per-component entropy gap $\to 0$

The key is “per-component entropy gap” means **entropy difference** / n.

In fact, the entropy gap is $O(\log n) \to \infty$. **Why is n the “right” scaling?**

Discussion of "The Shrinkage-Delinkage Trade-off"

Restricted variational families (mean field) can lead to poor posterior approximations.

Two very common responses to this problem are:

- "Machine learning": Ignore it (evaluate using some other criteria, like prediction)
- "Modern VI": Use more expressive families (at a computational cost)

This paper tries to understand *how* the restricted family goes wrong. *fire emoji*

This paper's most (initially) surprising conclusion is probably this:

Theorem 3.6, paraphrased

Let the target distribution have the constant ε-correlation matrix.
As the dimension n of the matrix goes to infinity:

- Each marginal mean field variance is wrong by ε
- The per-component entropy gap $\rightarrow 0$

The key is "per-component entropy gap" means **entropy difference** / n.

In fact, the entropy gap is $O(\log n) \rightarrow \infty$. **Why is n the "right" scaling?**

Why do the relative values across dimensions of the entropy gap matter?

It's clear why variance is useful. Less so the entropy gap, especially as n changes.